Haaga-Helia
ammattikorkeakoulu Oy

Tietovarastossa olevan tiedon laatu ja sen käsittely

Ari Helin

Opinnäytetyö
TIKO:n koulutusohjelma
2015

Tekijä(t)
Ari Helin

Koulutusohjelma
Tietojen käsittelyn koulutusohjelma

Opinnäytetyön otsikko

Tietovarastossa oleva tiedon laatu ja tiedon käsittely

Sivu- ja liitesivumäärä
6 + 2

Tämän opinnäytetyön tarkoituksena on selvittää yrityksen tietovaraston rakentamista.

Jatkuvasti kehittyvä tietovarasto avaa uusia ja monipuolisia mahdollisuuksia käyttäjien tietojen käsittelyyn, joten on mielenkiintoista tutkia tätä.

Tutkimuksella on kaksi pääaihetta: tiedon laatu sekä tiedon käsittely.

Tiedon laatu vaatii jatkuvaa ja systemaattista kehittämistä. Tiedon hallintaa varten tietoa pitää mitata. Tietovarasto, tiedon laatu ja eheys varmistetaan ETL-prosessin avulla. Tavoitteena on saada laadukasta tietoa tietovarastoon.

Käyttäjät odottavat oikeata tietoa tietovarastosta. Tietovarastossa on paras tieto, joka on saatu ja yhdistetty eri lähdejärjestelmistä.

Jatkotutkimuksena voisi selvittää, miten yrityksen tiedonhallintapolitiikka voitaisiin luoda ja kuinka ne pystyttäisiin jalkauttamaan.

Asiasanat
Tiedon laatu, Tiedon käsittely, ETL-prosessi, Tiedon laadun parantaminen

Author
Ari Helin

Degree programme
Business Information Technology

The title of thesis

The quality of the data warehouse and data processing

Number of pages and appendices
6 + 2

The purpose of this study is to investigate the construction of the company's data warehouse.

The study has two main themes: quality of data and data processing.

The quality of information requires a continuous and systematic development. Data for the management of information needs to be measured.

Data Warehouse, Data Quality and integrity is assured by the ETL process. The aim is to get quality information from the data warehouse.

Users expect correct data from the data warehouse. The best data is from the data warehouse that has been obtained and combined in different source systems.

Further research could examine how the company's data management policy could be created.

Key words
Data Quality, Data Processing, ETL process, Improving Data Quality

Sisällys

1 Johdanto

Opinnäytetyö on tehty omasta kiinnostuksesta johtuen. Olen nykyisessä työpaikassani ollut vetämässä tietovaraston kehitysprojektia. Projektissa olen todennut kuinka tärkeätä on tietovaraston tiedon laatu ja sen käsittely. Ne ovat tietovaraston olemassaolon edellytyksiä.

Tietovaraston hyödyntämiselle on vakiintuneita ja perusteltuja syitä. Tietovaraston avulla voidaan varmistaa yrityksen tietojen yksikäsitteisyys, yhteensopivuus ja yleensä tiedon laadun täyttyminen raportoinnin näkökulmasta. Sen avulla voidaan lisäksi yksinkertaistaa, ryhmitellä, summata ja esivalmistella tiedot raportointikäyttöä varten.

Tietovarastot myös ylläpitävät historia- ja versiotietoja, joita ei ole enää muilla keinoin saatavissa operatiivisista järjestelmistä. Tietovarasto vähentää operatiivisten järjestelmien kuormitusta ja kerää tiedot niistä erilliseen tietovarastoon ja sijoittaa tämän eri palvelimelle.

Tietovarasto on joukko eri prosesseja, kuten esim. tiedon lataus ja muokkaus sekä historiointi. Se sisältää valtavan määrä tietoa ja erilaisia laskentasääntöjä. Tietovarasto sisältää myös sovelluksia, laitteistoja, palvelimia ja henkilöitä, joilla on erilaisia rooleja. Usein puhutaankin tietovarastoympäristöstä.

Tässä opinnäytetyössä kuvataan mitä tietovaraston rakentamisessa on huomioitava tietovarastoon tuotavasta tiedosta ja sen käsittelystä.

Tutkimusongelma on tiedon laatu ja sen käsittely. Kuvaan opinnäytetyössä mitä tiedon laatu ja sen käsittely tarkoittaa. Esitän myös parannusehdotuksia tiedon laadun ja sen käsittelyn parantamiseen. Tiedon käsittelyllä tarkoitan tässä tapauksessa tietovarastossa käytettävää ETL-prosessia.

Tietovarasto sisältää tässä käsitteenä tietovarastoinnin ja raportoinnin. Datan tilalla käytetään sanaa tieto.

Yrityksen tietovarasto muuttuu jatkuvasti eri tarpeiden ja uusien lähdejärjestelmien takia, joten sitä on kehitettävä jatkuvasti.

Tässä opinnäytetyössä ei kerrota tietovaraston:

- Arkkitehtuurityypeistä

- Toteutusta
- Sovellusalustaa
- Sovelluksia
- Testausta
- Käyttöönottoa

Toisessa luvussa on aiheena liiketoiminnan hallinta (BI), jossa lyhyesti kerrotaan läpi mitä se tarkoittaa. Kerrotaan mikä on yrityksen tietovarasto ja miten se suunnitellaan raportointi huomioon ottaen. Lisäksi kerrotaan tietovaraston kehittämisestä, menetelmistä ja integroinnista. Tietovaraston elinkaaresta myös kerrotaan, mitä elinkaari tarkoittaa ja mitä siinä on huomioitava.

Kolmannessa luvussa on aiheena tietovarastoon tuotava tieto ja sen käsittely, jossa kuvataan tiedon laatua ja sen käsittelyä. Lisäksi kerrotaan miten tietovaraston tiedon laatua ja sen käsittelyä voidaan parantaa.

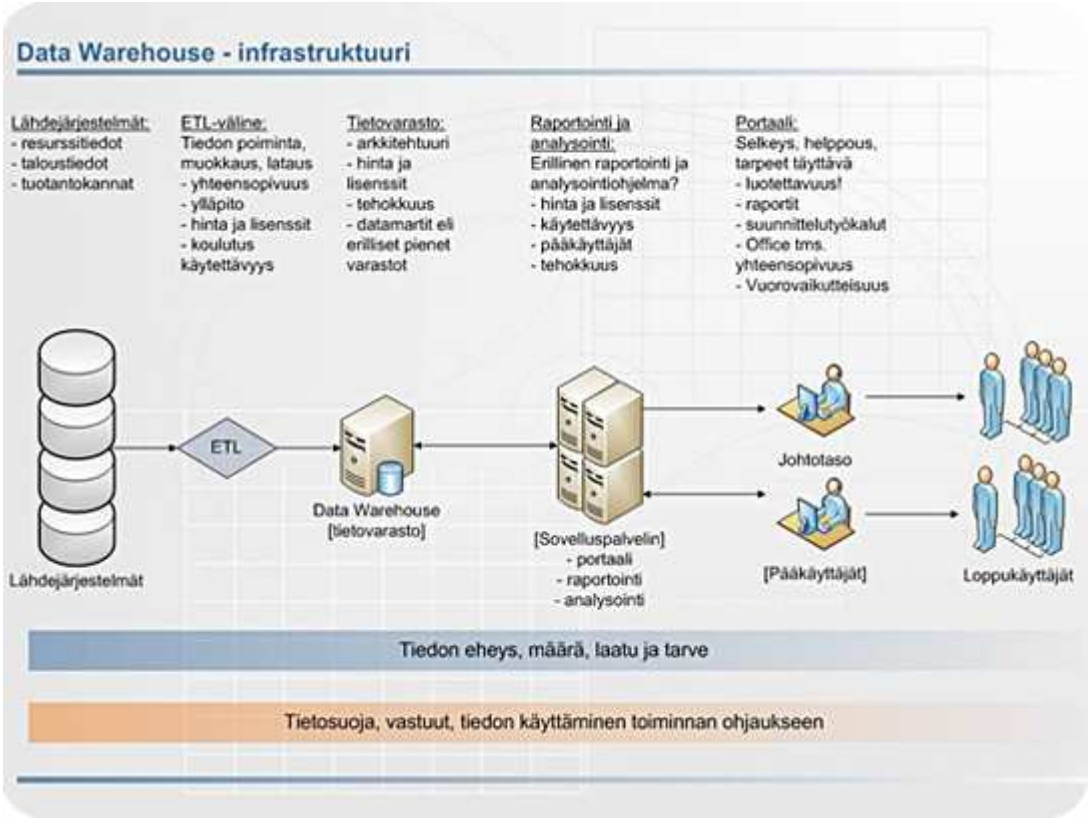
Neljännessä luvussa on aiheena johtopäätökset ja yhteenveto, jossa vedetään yhteen ja tiivistetään edellisten aihealueiden asiat sekä esitetään aiheiden käsittelyssä esiin tulleita johtopäätöksiä.

2 Liiketoimintatiedon hallinta

2.1 Yrityksen tietovarasto

Tiedolla olisi käyttäjiä! Kuitenkin monilla organisaatioilla on edelleen ongelmia saada tietoja edes omaan käyttöönsä analysoitavaksi, nopeassa, tuoreena ja yhdisteltynä. Tietojohtamisen tukeminen on vaikeaa, kun tiedot ovat hajallaan monien erillisten järjestelmien tietokannoissa. (Ari Hovi Oy 2012.) Tämä on yleensä operatiivisten järjestelmien ongelma.

Tietovarasto (DW) tarjoaa tähän ongelmaan ratkaisuvaihtoehdon. Tietovarastoon poimitaan ja integroidaan eri lähdejärjestelmistä tietoja yhteen. Tietovarasto on oma erillinen tietokantansa, joka tarjoaa monipuolisen ja entistä helpomman tavan tietojen raportointiin. Tietovaraston lataus tapahtuu yleensä joka päivä automaattisesti. Näin tiedot saadaan omiin käsiin jalostettuina ja riippuvuus tietojärjestelmätoimittajista vähenee. Keskitetyn tietovaraston tietoja voi hyödyntää sekä organisaation sisällä tai sen ulkopuolella. (Ari Hovi Oy 2012.) Tiedot ovat yhdestä paikasta eli tietovarastosta saatavissa.



Kuva 1. Business Intelligence, Data Warehouse ja ETL (Storberg, P. 2015).

Tietovaraston kokonaisarkkitehtuuri on yrityksissä usein yllä kuvatun kaltainen. Yrityksellä on tietotarve, johon suunnitellaan ja mallinnetaan tietovarasto. Siihen tuodaan tiedot eri lähdejärjestelmistä. Tietovaraston mallinnuksen jälkeen alkaa lähdejärjestelmien tietojen poiminta ETL-prosessissa.

ETL-prosessissa tieto puhdistetaan tyhjistä, päällekkäisistä ja vääristä tiedoista sekä samankaltaiset tiedot yhdistetään. ETL-prosessin loppuvaiheessa puhdistettu tieto siirretään yleensä latausalueelle myöhempää muokkausta varten ja siitä edelleen tietovarastoon sekä paikallisvarastoihin. Tietovarastossa tietoja voidaan käyttää hyödyksi raportointivälineellä.

Tietovarastossa tietoja käytettäessä virheet huomataan ja korjataan lähdejärjestelmissä. Näin tiedon laatukin paranee sekä tietovarastossa että lähdejärjestelmissä.

Tietovarasto on yleensä tietokanta, johon kerätään tietoa eri lähdejärjestelmistä. Näitä lähteitä ovat esimerkiksi sisäiset ja ulkoiset tietokannat ja tiedostot. Isoissa tietovarastoissa voi olla yli sata tietolähdettä.

Suurin osa yrityksen talletetusta kiinteämuotoisesta tietoresurssista sijaitsee operatiivisten järjestelmien tietokannoissa. Nämä järjestelmät palvelevat yleensä hyvin operatiivista toimintaa. Tietojen analysointia, raportointia ja satunnaisia kyselyjä sen sijaan on usein hidasta ja vaikeaa tehdä. Tiedot voivat myös sijaita eri järjestelmissä, maantieteellisesti hajautettuna. Tiedot saattavat sisältää paljon koodeja ja teknisiä vipuja. Kyselyihin ja raportteihin taas tarvitaan yksiselitteistä ja selväkielisempää tietoa. Operatiivisissa kannoissa ei myöskään yleensä voida tallettaa kovin paljon historiaa. (Ari Hovi Oy 2008.) Tietovarastossa olevan tiedon historiointi on yksi syy miksi tietovarastoja rakennetaan. Tietoa voidaan myös säilyttää tietovarastossa ”ikuisesti”.

Tietovaraston ideana on harmonisoida eri lähteistä kerätyt tiedot yhteismitalliseksi ja yhteen paikkaan, josta ne ovat kaikkien saatavilla. Tietovarasto on avoimesti käytettävissä halutuille tahoille, koska se on sijoitettu tietokantaan. (Louhia Consulting Oy 2015.) Tässä yhteydessä on kuitenkin huomioitava mahdolliset salassa pidettävät tiedot ennen kuin niitä voidaan jakaa kaikille halukkaille tiedon käyttäjille.

Keskitettyyn tietovarastoon kootaan mahdollisimman kattavasti kaikki se tieto, jota operatiivisessa, taktisessa ja strategisessa päätöksenteossa tarvitaan tänään ja huomenna (Midagon Oy 2011). Tietovarastoon pyritäänkin tuomaan kaikki mahdollinen tieto lukuun ottamatta tauluja, joissa on teknistä tietoa.

Tietovarasto toimii raportoinnin ja analytiikan tietolähteenä, josta eri raportointivälineet hakevat tietoa. Lisäksi tietovarastosta voidaan viedä tietoa myös takaisin operatiivisiin lähdejärjestelmiin. (Louhia Consulting Oy 2015.) Raportointi myös paljastaa lähdejärjestelmien tietojen laatu-ongelmat.

Tietovarastoon kannattaa tallentaa tulevaisuuden varalta sellaistakin tapahtumatason tietoa, joka on tällä hetkellä edullisesti käytettävissä ja jolla voi olla tulevaisuudessa liiketoiminnalle arvoa. Esimerkkinä voidaan pitää lääkärin potilaalle sähköisesti kirjoittamat reseptit ja niiden aikaleimat. Kansalainen voi tietojen avulla seurata tietovaraston avulla hänelle kirjoitettuja reseptejä.

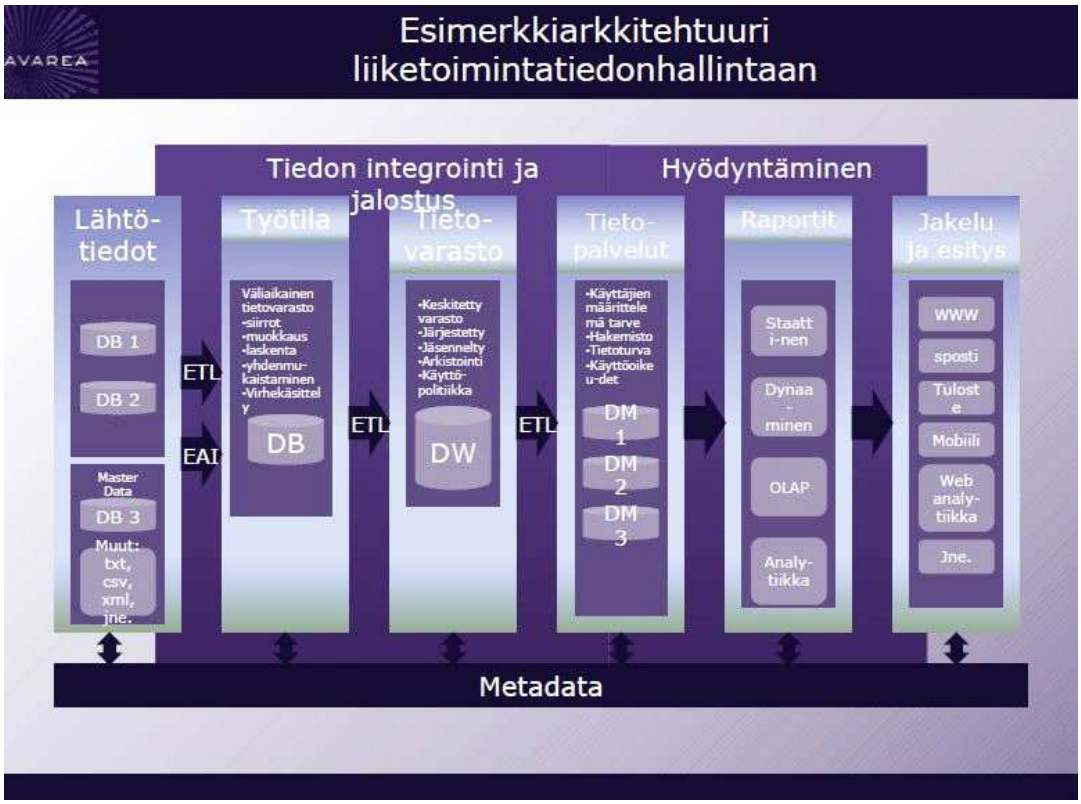
Keskitetyn tietovaraston avulla tiedon tarvitsijalla on vain yksi tietolähde (Midagon Oy 2011). Kaikki tarvittava tieto on yhdessä paikassa ja se on luotettavasti ja nopeasti hyödynnettävissä yhdestä paikasta.

Keskitetty tietovarasto kuulostaa ratkaisuna järkevältä, helpolta ja yksinkertaiselta (Midagon Oy 2011). Mutta sitä ennen on tietovarasto pystytettävä, arkkitehtuuri standardoitua, tieto- ja prosessivirheet minimoitava ja organisaation rakennettu tiedon jalostamisen kulttuuri.

Ideaalitilanteessa liiketoiminnalle tarjotaan koska tahansa ja mihin tahansa kysymykseen luotettava vastaus, niin tänään kuin huomennakin (Midagon Oy 2011). Tämä on tietovaraston yksi suurimmista tavoitteista.

2.2 Tietovaraston suunnittelu ja toteutus

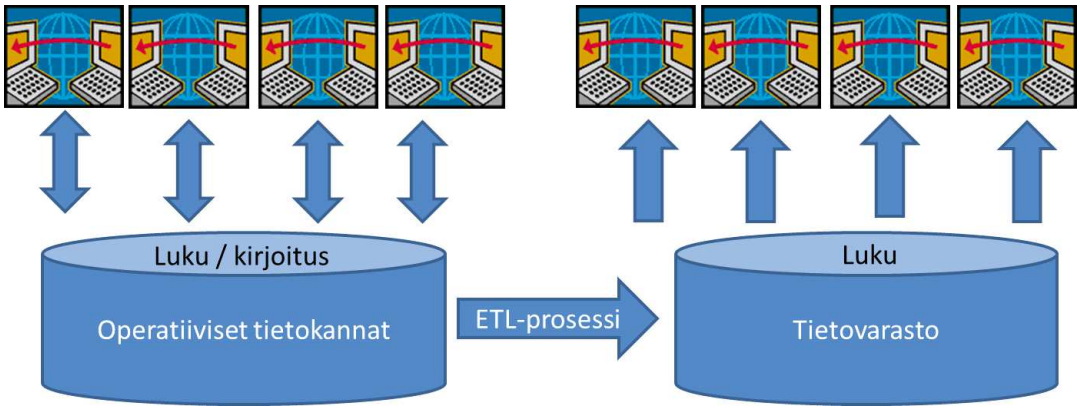
Tietovaraston arkkitehtuurityyppejä on monia. Yleinen tietovaraston arkkitehtuuri sisältää työtilan, tietovaraston (DW) ja paikallisvarastot (DM).



Kuva 2. Esimerkkiarkkitehtuuri liiketoimintatiedonhallintaan (Oksanen, M. 12.04.2011, s. 11).

Tietovarastointihanke koostuu yleensä seuraavista osa-alueista: tietovarastokannan mallintaminen, tietovarastolatausten suunnittelu ja toteutus sekä raportointi (Miracle Finland Oy 2015a).

2.3 Tietovaraston ja operatiivisen tietokannan erot



Kuva 3. Tietovaraston ja operatiivisen tietokannan erot

Tärkein ero tietovaraston ja operatiivisen tietokannan ero on yllä olevan kuvan mukaisesti se, että tietovarastoon ei kirjoiteta tietoa vaan ne ladataan eri lähdejärjestelmistä sinne.

Operatiivisen tietokannan tunnuspiirteitä

- Paljon lyhyitä päivittäviä tapahtumia
- Päivittävien tapahtumien tehokkuus tärkeää, päivitetty tieto pyritään saamaan mahdollisimman nopeasti myös muiden järjestelmien käyttöön
- Ajantasaisuus
- Ei historiointia

Tietovarastolle tyypillisiä tunnuspiirteitä

- Vain lukukäyttö
- Hakutehokkuus
- Päivitystehokkuus, ei-kriittinen paitsi ETL-vaiheessa joka vaikuttaa prosessin kestoon
- Tiedon historiointi, tietojen edelliset versiot mukana
- Tietokannan suuri koko historioinnin seurauksena
- Päivällä käytettävät tiedot ladataan varastoon yöllä.

2.4 Raportointi

Tietovarasto siis toimii raportoinnin ja analytiikan keskitettynä tietolähteenä. Erilaisten raporttien ja analyysien tekeminen on sieltä yleensä helppoa ja nopeaa, mutta aina jotakin

haasteitakin mukaan mahtuu. (Louhia Consulting Oy 2015.) Raportointi on siis tietovarastotyyppistä raportointia.

Tarvittava tiedon hyödyntämisen ratkaisu vaihtelee tietotarpeen ja käyttäjäryhmän mukaan. Tietovaraston tietoja voidaan hyödyntää joko laatimalla raportteja ja työpöytiä sekä tekemällä kyselyitä. Raportointi voi olla joko säännöllistä vakioraportointia (staattinen tai dynaaminen raportointi) tai Ad hoc -tyyppistä, tapauskohtaista raportointia.

Yrityksen sisäisessä käytössä puhutaan suorakäytöstä, joka voi tarkoittaa sitä että tiedon hyödyntäjä hakee itse tiedot kyselyitä käyttämällä tietokannasta tai tiedot tilataan IT-yksiköstä ja siellä suoritetaan tiedon kysely. Pääsääntöisesti käyttö kohdistuu tietovaraston paikallisvarasto-alueisiin, haastavampiin tietotarpeisiin käytetään tietovaraston DW-alueen tietoja. Poikkeustapauksissa käyttö voi kohdistua myös työtila-alueen tai lähdejärjestelmien tietoihin. Valvontatyyppiset tietotarpeet arvioidaan yleensä yrityksessä tapauskohtaisesti.

2.5 Iteratiivinen tietovaraston ja raporttien yhtäaikainen kehitys

Tietovaraston ja raporttien kehitys tapahtuu yhtäaikaisesti ja taikasanana mukana on iteratiivisuus (Eximia Business Intelligence Oy 2014). Iteratiivinen tietovaraston ja raporttien yhtäaikainen kehitys on lähestymistavoista yleisin. Se on osoittautunut käytännössä erilaisista lähestymistavoista parhaiten toimivaksi.

Iteratiivisuus antaa mahdollisuuden rakentaa raportteja tarvelähtöisesti. Oikein tunnistettu asiakastarve on edellytys sille, että raportin sisältö tuottaa eniten hyötyä käyttäjille. Raportteja ja tietovarastoa yhtä aikaa kehitettäessä käyttäjät saavat hyvin nopeasti konkreettista sisältö käsiinsä. He pystyvät siten nopeasti vaikuttamaan siihen, että raportti vastaa heidän tarpeitaan. Voidaan tuottaa esittämistä varten käyttäjille demoversio raportista.

2.6 Tietovaraston menetelmät

Kimball tietovaraston menetelmät kehitti Ralph Kimball, joka pidetään laajalti tietovaraston isänä. Suuntaviivat, joita jokaisen Kimball-tietovaraston olisi noudatettava ovat:

- Tietovaraston päätavoitteena pitäisi olla suorituskyky ja helppokäyttöisyys.
- Dimensionaalisia malleja voidaan kehittää ainoastaan kun tietoja koskevat vaatimukset on ymmärretty ja sovittu.
- Vaikka tietovarasto jatkuvasti kehittyvät, jokaisen projektin elinkaaren iteraation pitäisi koostua ennustettavasta toiminnasta, jolla on rajallinen alku ja loppu.

(Theta 2015.)

Kimballin tietovaraston menetelmää käytetään paljon pienissä ja keskisuurissa yrityksissä.

2.7 Tietovarastot ja integrointi

Tiedon analysointia ja kattavaa raportointia varten tarvitaan yrityksen eri perusjärjestelmissä olevat tiedot sekä mahdollisesti ulkopuolisista lähteistä haettavat tiedot yhteiseen tietovarastoon.

Tiedonsiirron suunnittelu

- Määritellään raportoinnissa ja analysoinnissa tarvittavat tiedot ja niiden lähteet
- Tarkistetaan tietojen yhteismitallisuus ja muodostetaan tarvittaessa tietokenttien vastaavuustaulukot
- Suunnitellaan tiedonsiirto-menettelyt toteutus tietovarastoon

(Talmax Oy 2013).

2.8 Tiedon elinkaari

Käsitteeseen tietovarasto liitetään hyvin helposti myös varaston sisältö eli itse tiedot. Kun ajatellaan tietovaraston elinkaarta, se on hyvin harvoin sama kuin tiedon elinkaari. Tietovaraston rakenne suunnitellaan ja varasto pystytetään. Useimmiten tietovaraston sisältö saadaan pääosin konversion tuloksena jostain vanhasta tietovarastosta, jonka rakenne ei enää vastaa sisällöllisiä tai teknisiä vaatimuksia. (Octel Oy 2008.) Vanha tietovarasto käy tarpeettomaksi ja se poistetaan käytöstä, mutta sen sisältämät tiedot ovat edelleen olemassa. Yrityksellä pitää olla suunnitelma miten toimitaan tällaisissa tapauksissa.

Tietovaraston elinkaari alkaa sen fyysisestä määrittelystä. Sen sisältö syntyy siis joko massana konversiosta tai vähitellen sovellusten erilaisten prosessien tuloksena. Koko elinkaarensa ajan tietovarastoa kohtaan kohdistuu kaksi vaatimusta, joihin sovellukset eivät vaikuta. Niiden mukaan tietovaraston pitää olla ehyt ja tehokas. (Octel Oy 2008.) Käyttäjät pettyvät tietovarastoon, mikäli sen tiedot eivät ole ehyitä ja se ei toimi tehokkaasti. Silloin tietovaraston käyttö jää vähäiseksi.

3 Tietovarastoon tuotava tieto ja sen käsittely

Tietovaraston tiedoissa ja niiden käsittelyssä on monia seikkoja mitkä pitää huomioida. Tietovarastoon tuodaan tietoja eri lähdejärjestelmistä ja niiden tietoja yhdistetään ja puhdistetaan tiedon käsittelyssä (ETL-prosessissa).

3.1 Käyttäjät vaativat tietoja

Käyttäjät ja johto yrityksissä ovat usein turhautunutta tiedon vaikeaan ulossaantiin järjestelmistä. Tiedetään, että tarvittava tieto on olemassa, mutta yhteenvetojen ja pikaraporttien saanti on niin hidasta, työlästä ja kallista, että on luovuttu toivosta. (Ari Hovi Oy 2008.)

Käyttäjät tarvitsevat nopeita yhteenvetoja sekä summatietoja ja tiettyyn tarkoitukseen tehtyjä raportteja. Isoissa yrityksissä voi olla monia hajautettuja operatiivisia järjestelmiä. Näiden eri järjestelmien tietokannat eivät yleensä ole saman toimittajan vaan voi olla monia eri tuotteita käytössä. Vaikka tietokannat olisivat samaa tuotetta on vaikeaa tehdä yhteenvetoraportteja erilaisista systeemeistä.

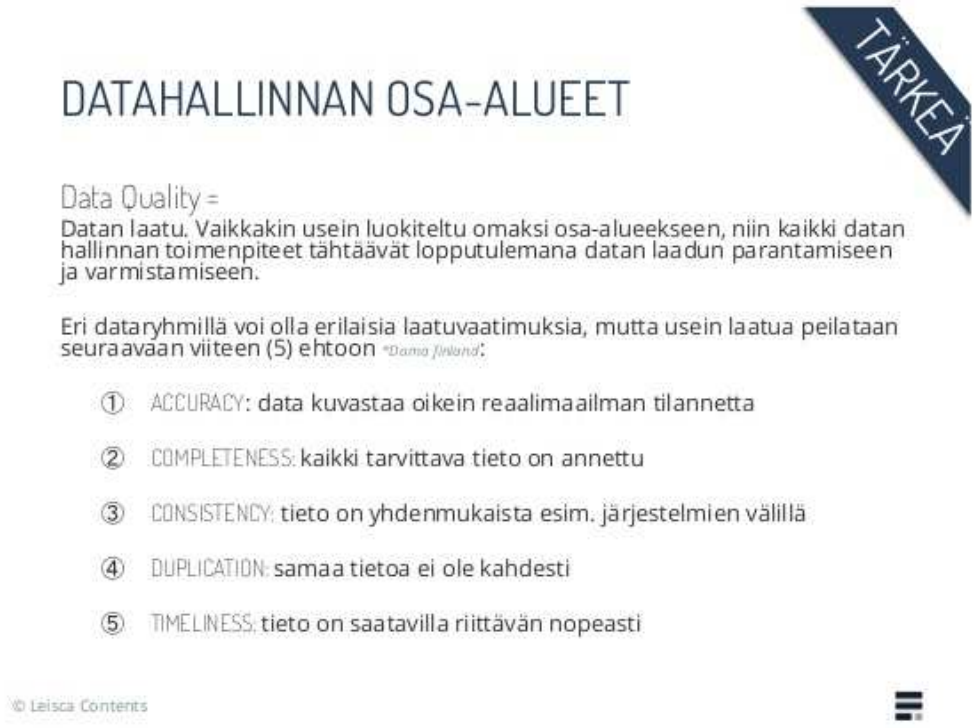
Operatiiviset kannat on suunniteltu tehokkaaseen tapahtumankäsittelyyn, joka on usein päivityspainotteista (Ari Hovi Oy 2008). Niissä tietojen toistoa ja summaamista vältetään päivitysten hidastumisen takia. Tietokannat on hyvin normalisoitu ja ne ovat usein raskaassa tapahtumakäytössä. Käyttäjät eivät yleensä saa tehdä kyselyjä operatiivisiin kantoihin, koska seurauksena on operatiivisen toiminnan huomattava hidastuminen. Päiväsaikaan tehtäviä kyselyjä ja laajasti kantaa selaavia raporttiajoja eivät operatiiviset tietokannat kestä. Ne eivät usein myöskään olemassa olevalla rakenteellaan tue helppoja kyselyä ja raportointia.

Ratkaisu edellä kuvattuihin ongelmiin on tietovarasto (Ari Hovi Oy 2008). Tiedot poimitaan operatiivisista lähdejärjestelmistä, jotka sitten ladataan erilliseen tietovarasto-kantaan. Tämä tietovarastokanta on suunniteltu juuri kyselyjä ajatellen. Tietovarastokannassa tietoja toistetaan. Tällä tavalla saadaan taulujen määrä pienemmäksi ja samalla tarvittavien liitosten määrä pienenee. Tämän seurauksena kyselyjen suorituskyky on huomattavasti parempi kuin täysin normalisoidun kannan. Kiinnitetään siis huomiota raporttien toteuttamisen nopeuttamiseen ja ajettavien kyselyjen suorituskykyyn.

Operatiivisten sovellusten raportointipuoli saattaa jäädä tässä arkkitehtuurissa huomattavasti pienemmäksi kuin aiemmin, mikä voi parantaa sovelluskehityksen tuottavuutta.

3.2 Tiedon laatu

Kaikki tiedon hallinnan toimenpiteiden tarkoituksena on tiedon laadun parantaminen ja varmistaminen alla olevan kuvan mukaisesti. Dama Finlancin mukaan laatuvaatimuksia on viisi.



Kuva 4. Data-suomi-sanakirja (Niemi, K. 18.4.2013).

Seuraavaan ja ylivoimaisesti hankalimpaan ongelmaan törmätään, kun tiedetään, mistä tarvittavat tiedot on saatavilla, ja kuinka ne sieltä saadaan. Ongelman nimi on tiedon luotatettavuus ja laatu. (Midagon Oy 2011.) Tietovarasto paljastaa armotta lähdejärjestelmien tiedon rakenteellisen ja sisällöllisen laadun puutteet. Esimerkiksi lisätty vahingossa reseptiin lääkkeen, jota ei enää ole olemassa. Tai lisätään potilaan vanha osoite hyvässä uskossa.

Haastavimpia tiedon laatuvirheitä ovat tiedon sisältövirheet. Esimerkiksi potilaan lääkkeen hinnaksi on tallennettu väärä hinta. Tällaisia virheitä ei voi tietovarastossa saada automaattisesti kiinni.

Laadukkaan tietovaraston takana on hyvä malli (InfoBuild Oy 2010b). Tietojen mallintaminen onkin yksi tärkeimmistä tiedon laatuun vaikuttavista tekijöistä.

Tieto pitää ensin mallintaa, ennen kuin sitä voidaan mitata tai varastoida. Eräajon kaatumisen syynä saattaa olla huonolaatuinen tieto. Silloin pitää selvittää tiedon alkuperä. (InfoBuild Oy 2010b.) Raportoinnissa selviääkin usein lähdejärjestelmien tietojen laatuun liittyvät ongelmat.

Lähdejärjestelmän kokonaiskuvan hallitsemiseksi on tärkeää tietää missä alkuperäinen tieto sijaitsee. Tiedon alkuperän mallintaminen kertoo mistä tieto on alun perin kotoisin ja missä kyseisen tiedon master tietoja pitää hallita.

Sama ongelmatiikka on tietovarastoinnissa ja tiedon laadun mittaamisessa (InfoBuild Oy 2010b). Tiedon laatua pitääkin mitata aina säännöllisesti.

Tiedot yleensä siirretään työtila (staging) -alueelle, jossa tieto puhdistetaan ja yhtenäistetään. Työtila-alueella tietoa säilytetään vain vähän aikaa.

Seuraava kompastuskivi on eri järjestelmistä saatavien tietojen yhdistäminen ja yhdenmukaistaminen (Midagon Oy 2011). Lähdejärjestelmissä on suuri määrä erilaisia avaintietoja, joita käyttäjät eivät näe, eivätkä tiedä ja eikä tarvitsekaan välittää. Eri lähdejärjestelmien tietojen yhdistämisessä tarvitaan kuitenkin näitä avaintietoja. Potilaiden yhdistävä avaintieto on henkilötunnus. Mutta joissakin tapauksissa esiintyy pelkkä syntymäaika.

Selvitettävä minkä tiedon pohjalta eri järjestelmissä olevat tiedot voidaan yhdistää toisiinsa, jos henkilötunnuksia tai syntymäaikoja ei voida käyttää. Virheellinen tai puuttuva linkitys eri lähdejärjestelmien välillä aiheuttaa tiedon eheysvirheitä tietovarastoon.

Tietovarastossa vaaditaan tietojen jäljitettävyyttä. Tietovaraston käyttäjät haluavat tietää tietojen alkuperän, tietojen mahdolliset muokkaukset, tietojen tuoreuden ja päivittämättä jääneet tiedot.

On selvitettävä, mistä järjestelmästä tai järjestelmistä tarvittavat tiedot löytyvät, missä muodossa ne siellä ovat, ja missä muodossa ja millä keinolla ne ovat sieltä saatavissa (Midagon Oy 2011). Samat tiedot saattavat löytyä monista eri järjestelmistä ja niiden tietoja voidaan päivittää. Esimerkiksi on asiakastietojen ylläpitäminen SAP-järjestelmässä. Selvitettävä minkä järjestelmän tiedot ovat luotettavimpia ja ajantasaisimpia. Selvitettävä myös onko tietoa saatavilla jostain järjestelmästä vai voidaanko se yhdistellä tai päätellä muiden tietojen avulla. Selvityksen perusteella tietoa voidaan ostaa myös organisaation ulkopuolelta.

Tietovarasto-kantaan monasti summataan valmiiksi usein kysyttäviä tietoja, kuten esim. reseptien lukumäärä kuukausi, potilaiden määrät tai reseptien lukumäärät vuoden alusta. Summataulut ovat huomattavasti pienempiä kuin tapahtumataulut, mikä nopeuttama kyselyjen suorituskykyä. Kyselyjen tekeminen myös helpottuu.

Tietojen toistamisesta herää kysymys, että eikö tiedon eheys ole vaarassa. Sama tieto eri paikoissa voi mennä solmuun. Tietovarasto-kannan tiedot ladataan ja päivitetään keskitetysti yhdestä pisteestä. Tässä otetaan huomioon jo kaikki tiedon toistamiset ja tehdään tarvittavat summaukset. Suorapäivityksiä ei voi sallia. Vaara eheyden menettämiselle on siis hyvin pieni.

Selvitettävä miten tietojen jäljitettävyys ratkaistaan yrityksessä. Tietovaraston ylläpito myös vaatii, että eri lähdejärjestelmistä tulleet tiedot latausaikatauluineen on kirjattu ylös ja jälkikäteen tutkittavissa. Tietojen jäljitettävyys vaaditaan, kun tietoja summataan ja yhdistellään. Kuinka voidaan todentaa, mitkä lähdejärjestelmistä saadut tapahtumat sisältyvät esimerkiksi tiettyyn tuntitasolle laskettuun tilastoon.

Kokemus on osoittanut, että tiedon jakaminen ja käyttöoikeuksien hallinnointi yhdessä keskitetyssä paikassa on helpompaa kuin tietoturvan ylläpito IT-spagetissa (Midagon Oy 2011). Selvitettävä tietovaraston omistajuus, jotta voidaan hyödyntää siitä saatavaa tietoa optimaalisesti.

Tietojen ja järjestelmien omistajuus ei ole yrityksissä vakiintunut käsite tietojen omistajuuden osalta. Perinteisesti aiheeseen ei ole liittynyt problematiikkaa, koska useimmissa tapauksissa järjestelmän ja tiedon omistajana on ollut sama tulosyksikkö. Se mitä omistaja on päättänyt järjestelmän suhteen, on käytännössä ohjannut myös tiedonhallinnan ratkaisuja. Toimintaa on leimannut tietty järjestelmäkeskeisyys.

Kehitys vie kuitenkin kohti palvelukeskeistä kokonaisarkkitehtuuria, jossa järjestelmien ja tietojen omistajuus eivät automaattisesti kohtaakaan. Tällaisessa toimintaympäristön muutoksessa myös omistajuuden käsitettä on tarkistettava ja se tarkoittaa päätöksentekoroolin laajentamista aktiivisen ohjausvastuun sekä yhteistoimintavelvoitteen suuntaan. Kyseessä on kulttuurinen muutos. Se ei etene itsestään, vaan sitä tulee tietoisesti edistää hanke- ja projektitasolla.

Käytettävyys on laatutekijä käyttäjän näkökulmasta (Teknologian tutkimuskeskus VTT Oy 2015). Tietovaraston käytettävyys määrittelee, ratkaiseeko tuote käyttäjän näkökulmasta tarpeet oikealla tiedolla ja ratkaiseeko tuote tarpeen käyttää sitä helpolla tavalla.

Käytettävyyteen vaikuttavia tekijöitä ovat:

- Tarkoituksen soveltuvuuden tunnistaminen
- Opittavuus
- Helppokäyttöisyys
- Käyttäjän virheiltä suojautuminen
- Käyttöliittymän esteettisyys
- Saavutettavuus
- Porautuminen tiedoissa alimmalle tapahtumatasolle mahdollista

Tehokkuus kuvaa, kuinka tehokkaana tietovarasto koetaan henkilöiden päivittäisessä työssä. Tehokkuuden kokemukseen vaikuttavat, kuinka helposti tuote on saatavissa, kuinka nopeasti tieto on käytettävissä, kuinka paljon käyttäjä kohtaavat virheitä tuotteen käytön yhteydessä ja kuinka helposti tuote on opittavissa

Opittavuuteen vaikuttaa se, että tietovaraston täytyy olla helposti tavoitettavissa. Raportin täytyy olla hyödyllinen, jotta käyttäjä saa mitä haluaa, ja on suunniteltu tuottamaan sitä tietoa, mitä käyttäjä tarvitsee. Raportilla ei saa olla liikaa valintoja, jotka pysäyttävät käyttäjän harkitsemaan, mitä minun pitää tehdä saadakseni tämän tiedon. Käyttäjälle tuotettava tieto täytyy olla käyttäjälle tuttua ja hyvin kuvattua, sekä pyrkiä käyttämään samanlaisia ratkaisuja eri käyttöliittymissä.

Tarkistetaan poikkeavatko kirjausmenettelyt ja kooditukset (esim. asiakasnumerot, projektikoodit jne.) eri järjestelmien välillä, Tarvittaessa rakennetaan kartta yhdistämään eri kooditukset. Kartan avulla saadaan tiedot siirrettyä tietovarastoon yhtenäistettyinä. (Talmax Oy 2013.)

Eheys pitää ymmärtää tässä varsin fyysisenä ja teknisenä asiana eikä sillä ole mitään tekemistä tietojen loogisen oikeellisuuden kanssa. Siitä vastaavat sovellusohjelmat. (Ari Hovi Oy 2008.) Tämä on siis tekniikkaa.

Tietovarasto ei ole ehyt, jos se sisältää esim. epätäydellisiä päivityksiä tai rivejä/segmenttejä joihin hakemisto/osoitin ei viittaa (Ari Hovi Oy 2008). Tämä kuuluu tietovaraston jatkuvaan seurantaan, jota ylläpito tekee.

Tiedonhallintajärjestelmä takaa sen, että tietovarasto on ehyt tai vaurion tapahduttua pystytään palauttamaan ehyeksi, jos tietovaraston hoitaja hoitaa oman osuutensa: tallettaa

lokit sekä ottaa ja tallettaa varmuuskopiot, eikä syyllisty omavaltaisuuksiin (Ari Hovi Oy 2008). Nämä ovat normaaleja tietokannan hoitotehtäviä.

Tehokkuus taas merkitsee sitä, että tietojen käsittely on niin nopeata kuin mahdollista eikä mitään resurssia tuhlata (Ari Hovi Oy 2008). Raportoinnissa tämä näkyy raportin ajon nopeutena, kuten ETL-ajoissakin.

Kun tiedot on ladattu tietovarastoon tai se on uudelleenjärjestetty, ollaan lähtötilanteessa (Ari Hovi Oy 2008). Yleensä tiedot ladataan tietovarastoon päivittäin.

Tiedon laatu rakentuu ihmisten, prosessien ja teknologian kautta (Innofactor Oyj 2013). Ihmisillä on suuri vaikutus tiedon laatuun niin tiedon tallennuksessa kuin sen seurannassakin.

Tiedon laatua ylläpidetään yrityksessä tiedon omistajuuden ja prosessin avulla. Näiden lisäksi tarvitaan automaattisia ja joustavia tiedonkäsittelytapoja laadunhallintaan. (Innofactor Oyj 2013.)

Yksi totuus tiedolle ja siihen liittyvät sovellukset ovat kovassa kasvuvauhdissa. Eikä ihme, onhan erilaista, tärkeää liiketoiminnan seurantaan ja ohjaukseen tarvittavaa tietoa monesti ripoteltuna ympäri organisaation Excel-viidakkoa. (Rongo Oy 2014.)

Mitä paremmin yrityksessä liikkuva tieto hallitaan, sitä todennäköisemmin sen avulla voidaan lisätä koko liiketoiminnan tuottavuutta (InfoBuild Oy 2010b). Tärkeätä on se, että ihmisille välitettävä tieto on visuaalista ja aidosti ymmärrystä lisäävää.

Tiedon omistajuuden näkökulma, eli se, kuka ja mistä tietoa hallitaan (InfoBuild Oy 2010b). Kenelläkään ei ole tarkkaa tietoa tiedon oikeellisuudesta tai ajantasaisuudesta eri lähdejärjestelmissä eikä siitä, missä järjestelmässä tieto on valideinta.

Keskeisintä tietovaraston toimivuudessa on sen johdonmukaisuus (InfoBuild Oy 2010b). Tietovaraston tietojen on oltava sisällöltään johdonmukaisia ja ymmärrettäviä. Tällöin termit ja sanasto ovat yritykselle tuttuja. Johdonmukaisuudella tarkoitetaan tässä lisäksi yhtenäisiä merkintöjä sekä määritelmiä tietovaraston eri tietolähdejärjestelmistä kootusta sisällöstä.

Tiedot siirretään ajoittain operatiivisista kannoista tietovarasto-kantaan. Tiedot eivät siis ole ajan tasalla. Siirtotiheys voi olla esim. kuukausi, viikko tai päivä. (Ari Hovi Oy 2008.) Tietojen analysointia, raportointia ja kyselyjä varten riittää yleensä päivän vanha tieto. Käyttäjät haluavat tietää tietojen ajantasaisuuden asteen.

Operatiivisten kantojen tiedot ovat usein koodattuja, jopa useita koodeja samassa kentässä. Eri järjestelmissä saattaa samalle asialle olla eri tietotyyppi. Pvm-muodot voivat olla erilaisia. Jopa käytetään eri tunnuksia (esim. asiakastunnus) eri järjestelmissä. (Ari Hovi Oy 2008.) Kaikki tämä tieto on puhdistettava tietovarasto-kantaan tietoja ladattaessa. Jos on virheellisiä tietoja, ne hylätään. Ylensä koodit puretaan teksteiksi raportointia varten. Lisäksi tehdään jo summauksia valmiiksi jatkokäyttöä ajatellen.

Tietovarasto -kantaan halutaan yleensä säilöä myös historiaa. Tämä tarkoittaa sitä, että kannan tietoja ei päivitetä, vaan kantaan aina lisätään uusia tietoja esim. kaudella tai pvm:lla varustettuna. Näin Tietovarasto-kantaan alkaa kertyä trendianalyysejä varten historiaa, jota käyttäjät jatkuvasti tarvitsevat liiketoimintaa analysoidessaan. Tietokannan koko samalla myös kasvaa. Kannan koko riippuukin ratkaisevasti siitä, paljonko historiaa halutaan säilyttää ja millä tarkkuus- tai summaustasolla.

Lähdejärjestelmän tekninen vastuu tarkoittaa neljää asiaa:

- Tieto on saatavilla silloin kuin Etl-prosessi sen tarvitsee käyttöönsä.
- Tiedon oikeellisuus, tieto on sitä mitä se väittää olevansa.
- Lähdejärjestelmän vastuulla on virheellisen tiedon korjaus.
- Halutaan saada jäljitettävyys ETL-prosessin eri vaiheisiin.

3.3 Ydin- ja metatietojen hallinta

Yrityksen operatiivisten järjestelmien ydintietojen (Master Data) laadusta huolehtiminen on tärkeä osa menestyksellistä liiketoimintaa.

Ydintieto (Master Data) on yrityksen toiminnalle välttämätöntä tietoa, kuten asiakkaiden ja toimittajienkin tiedot, joita käytetään useassa eri järjestelmässä. Yrityksissä, joiden ydintietojen laatu on kunnossa, prosessit toimivat kitkattomasti ja päätöksenteko on tehokasta. (Innofactor Oyj 2013.) Tämä on tietovarastoon ladattujen tietojenkin kannalta ydinkysymys.

Ydintietojen hallinnalla (Master Data Management, MDM) tarkoitetaan pääasiassa eitapahtumien lähtötietojen yhtenäistämistä organisaatiossa. Näin pystytään hallitsemaan useista lähteistä tulevaa, mahdollisesti samaa tai samankaltaista tietoa ja yhdistämään ne yhdeksi hallituksi tietolähteeksi ja hyödyntämään sitä mm. raportoinnissa.

Määritellyt prosessit hyödyntävät laadukasta tietoa ja näin saavutetaan monia liiketoiminnallisia hyötyjä. (Innofactor Oyj 2013.) Heikko tiedon laatu sekä niiden epäjohdonmukai-

suus ovat selviä merkkejä siitä, että yrityksen ydintiedot on tarpeen laittaa pikaisesti kuntoon.

Metatiedolla (metadata) tarkoitetaan tietoa tiedosta. Tietovarasto-ympäristössä metatiedolla ymmärretään tietovaraston ja raportoinnin välissä olevaa kerrosta, jossa kunkin yrityksen ymmärtämin termein selitetään tietovarastossa olevia käsitteitä. Metatieto-mallissa pyritään myös piilottamaan loppukäyttäjältä tietovaraston monimutkaisuutta. Mikäli tietovarastoon on päässyt livahtamaan esimerkiksi sinne kuulumattomia toimintoja, voidaan niitä suodattaa pois metatieto-kerroksessa. Tällöin ei tarvitse tehdä vastaavia muutoksia muille raporteille, jota sama muutos saattaisi koskea. Käänteisesti metatieto-malliin tehty ehkä yksinkertaiselta vaikuttava rakenteellinen muutos saattaa vaikuttaa muihin raportteihin kuin on suunniteltu.

Yrityksen hyvin suunnitellulla metatietomallilla ja kyselyiden optimoinnilla raportointiväline toimii paremmin loppukäyttäjien tarpeen mukaan. Haettavan tiedon relevanssia ja nopeutta voidaan näin parantaa merkittävästi. Tiedonhaun kehittäminen on loppukäyttäjien kannalta tärkeää. Pitää myös tiedon katselmoinnilla varmistaa, että tehdyt parannukset vastaavat hyvin käyttäjien tarpeita ja toiveita.

3.4 Tiedon jatkuva ylläpito

Jatkuvaan tiedon laadukkaaseen ylläpitämiseen tarvitaan selkeät määritykset, prosessit ja työkalut. Tiedolla tarkoitetaan yrityksen yhteistä tietoa. Tämä tieto liittyy yleensä potilaiden ja lääkäreiden kaltaiseen perustietoon. Jos tiedossa on virheitä tai puutteita, raportointia ja analytiikkaa ei saada todenmukaiseksi. Tiedon jatkuva ylläpito pitää huolen, että tieto pysyy oikeellisena ja ajantasaisena.

Tiedon jatkuvan ylläpidon kokonaisuus sisältää:

- Tiedon liiketoiminnallisten vaatimusten analysoinnin
- Olemassa olevan tietovaraston kartoituksen
- Tiedon ylläpidon prosessin suunnittelun liiketoiminnan tarpeita vastaavaksi
- Palvelukuvauksen ja palvelutasosopimuksen laatimisen toimittajan kanssa
- Asiantuntijatiimin ja henkilöstön koulutuksen
- Tiedonvälityksen järjestelmän käyttöönoton
- Jokapäiväisen tiedon sisällön päivitys- ja harmonisointioperaatiot
- Toteutettujen muutosten sekä tapahtumien raportoinnin
- Palvelun laadun seurannan ja jatkuvan kehittämisen

Tyypillisiä jokapäiväisiä tiedon ylläpidon toimenpiteitä ovat:

- Tietoon liittyvien päivityspyyntöjen hallinta
- Manuaalisten tiedon lähteiden käsittely, mikäli niitä esiintyy
- Nimikkeiden perustaminen ja päivittäminen
- Tiedon rikastaminen
- Tiedon harmonisointi
- Tiedon puhdistaminen ja duplikaattien poistaminen
- Tiedon laadun tarkistus ja varmistus
- Raportointijärjestelmän sisällön päivittäminen
- Virhetilanteiden käsittely
- Systemaattisia ja kurinalaisesti noudatettuja ylläpitoprosesseja.

Hyödyt

- Ydintieto pysyy aina kunnossa
- Antaa joustoa ja varmistaa laadun
- Säästää kustannuksia merkittävästi jo lyhyellä aikavälillä

Tietovarastoinnin tietoon liittyviä ongelmia ovat:

- Paljastuneet tiedon laadun ongelmat operatiivisissa tietokannoissa
- Vaaditun tiedon puuttuminen operatiivisista tietokannoista
- Tiedon homogenisointi
- Tiedon omistusoikeudet

3.5 Tietoa käsittelevä ETL-prosessi

ETL-prosessissa tietoja haetaan lähdejärjestelmästä ja ladataan tietovarastoon. Tietojen liikuttelu voi muodostua ongelmaksi, jos siirrettävää tietoa on paljon, mutta yhteydet ovat huonot. Esimerkiksi näin voi olla jossain päin Suomea.

Selvitettävä miten varmistetaan kansalaisen tietojen saatavuus ajoissa ja kattavasti. Selvitettävä myös tukkiiko eri järjestelmien yhtäaikainen käyttö organisaation verkon. Tai tietovaraston yksittäinen ETL-latausajo ei onnistu. Väliin jäänyt ETL-latausajo jättää informaation eheyteen aukkoja.

Tietovaraston muodostus tapahtuu ETL-prosessin kautta. ETL-prosessi vastaa lähdejärjestelmistä tulevan tiedon anonymisoinnista ja tallentamisesta työtila-alueelle sekä tiedon oikeellisuustarkistuksia**.**

ETL-prosessi vastaa myös työtila-alueella olevien tietojen yhtenäistämisestä tietovaraston tietomalliin. Tämän lisäksi siinä tehdään muutoksia tiedolle niiltä osin kuin on tarpeellista. Lisäksi tarvittava tiedon historiointi voidaan hoitaa.

ETL-prosessi vastaa myös tietovaraston tietojen saattamisesta paikallisvarastoihin. Tiedot viedään paikallisvarastoihin, jolloin niitä on helpompi käsitellä raportointi-, tilastointi- ja analytiikkaohjelmistoilla.

ETL-prosesseista tuotetaan raportteja, mm. onnistumisesta, tietomääristä ja prosessin kestosta. Mikäli ETL-prosessi epäonnistuu, asiasta lähtee viesti ylläpitoon.

Tyypilliset hälytyksiin liittyvät tiedotettavat tilanteet ovat:

- Aineiston ajantasaisuuden vaarantuminen
- Aineiston päivittyminen myöhässä
- Tiedossa havaittiin merkittävä poikkeama aineiston määrässä tai arvoissa
- Ajon kaatuminen vaatii toimenpiteitä

Tietovarasto on tietokantajärjestelmäkokonaisuus. Tietovaraston ytimessä on tietokanta, johon yhdistetään tietovarannot, joita halutaan käyttää yrityksessä oleviin raportointitarpeisiin. Tietovarannot jakautuvat tapahtumatietoihin (esim. potilasjärjestelmät) ja tapahtumatietoja yhtenäistäviin (master data) ja luokitteleviin (koodistot) tietoihin. Tapahtuma- ja master data-tietoja saadaan myös yrityksen ulkopuolelta.

Tiedon laatu rakentuu ihmisten, prosessien ja teknologian kautta. Tiedon laatua ylläpidetään organisaatiossa tiedon omistajuuden ja prosessin avulla. Näiden lisäksi tarvitaan automaattisia ja joustavia tiedonkäsittelytapoja laadunhallintaan. Tieto, jota ei jalosteta, jää vain raaka-aineeksi ja siinä hukataan paljon organisaation potentiaalia. (Innofactor Oyj 2013.) Tässä on tietovaraston tiedonkäsittelyn menetelmien läpinäkyvyys ja kommunikoitavuus erityisen tärkeää. Usein tietovarastoa hidastavaksi tekijäksi muodostuu tiedon jalostus ennen kuin tietoa päästään mitenkään hyödyntämään.

Tietovarastohanke jalostaa yrityksen operatiivisten järjestelmien tietoja myynnin, yritysjohdon ja koko yrityksen tarpeisiin. Tietovarastokantaan tiedot on ensin siirrettävä ts. ladattava operatiivisista lähteistä ja niitä on samalla yhteismitallistettava, puhdistettava ja jalostettava. Usein tietovarastohanke paljastaa puutteita ja laatuongelmia operatiivisissa tiedoissa, jolloin samalla parannetaan operatiivisten järjestelmien tietosisällön laatua. Toinen osa tietovarastohanketta on tiedonluovutus, jolloin tietovaraston tietoja hyödynnetään eri tavoin organisaatiossa ja sen ulkopuolella. (Miracle Finland Oy 2015.) Raportoinnin tietoja voivat käyttää yrityksen sisäisten käyttäjien lisäksi muut yksityiset ja julkiset organisaatiot sekä viranomaiset.

Kaikkea yrityksen päivittäisiä toimintoja ei kannata tallentaa tietovarastoon vaan latausvaiheen aikana siistitään tietoja ja poimitaan vain merkityksellisiä asioita. Tietoja ei tallen-

neta yhtä tarkalla tasolla kuin mitä lähdejärjestelmistä saattaisi löytyä vaan tiedot koostetaan ylemmälle tasolle. Tätä kutsutaan tiedon yksityiskohtaisuudeksi.

Myös päällekkäisyyksiä pyritään poistamaan ja yhdenmukaistamaan. Tietovarastossa on tärkeätä panostaa tietojen laatuun, jotta saavutettaisiin vain yksi totuus tiedosta.

Suunnittelutyö ja toteutus sisältää seuraavanlaisia tehtäviä:

- Standardit ja laatu, jossa sovitaan standardit ja laatuvaatimukset
- Toteutuksen laajuus, jossa rajataan kriittiset tiedot ja tiedonluovutustarpeet
- Määritellään vaatimukset latauksille, tiedonluovutukselle ja raportoinnille
- Tekninen arkkitehtuuri.

(Miracle Finland Oy 2015.)

3.6 Tiedon laadun analysointi lähdejärjestelmissä

Lähdejärjestelmien tiedon laadun analysoinnilla on merkittävä rooli tiedon laadun parantamisessa. Se tulisikin suorittaa jo ennen kehityksen ja testauksen aloitusta viimeistään määrittelyvaiheessa. Kun tiedon laatuun liittyvien riskien vaikutukset pystytään minimoimaan mahdollisimman aikaisin, pystytään välttymään yllätyksiltä myöhemmissä testausvaiheissa.

Tiedonlaadun analysoinnin avulla luodaan karkea peruskäsitys eli profiili tarkasteltavan tietolähteen sisällöstä (tiedostosta, tietokantataulusta, reaaliaikaisesta taustajärjestelmäliittymästä, sanomaliikenteestä jne.). Sen avulla saadaan nopeasti ja helposti käsitys siitä, miten tietyn tieto-elementin arvot ovat jakautuneet, mitkä ovat tyypillisimmät arvot, maksimit, minimit jne., samoin minkälaisia sisällön muotoja ja malleja tiedosta löytyy.

Tiedon laadun mittaamisessa keskeistä on aika. Yleensä halutaan, että virheet paljastuvat heti (reaali-ajassa tai lähes reaali-ajassa) ja toisaalta halutaan että pitkän aikavälin (esim. vuosien mittainen) kehitys on selvästi nähtävissä. (InfoBuild Oy 2010a.) Tiedon laatua mittaamalla löydetään kehittämiskohteita ja luodaan parempaa tiedon laatua.

Tiedon johtaminen vaatii hyvää tiedon laatua. Ydintiedon hallinnan tärkein ajuri onkin tiedon laatu. Pääasiallisia tiedon laatuongelmia ovat puutteet tiedon oleellisuudessa, ajankohtaisuudessa ja luotettavuudessa. Laatuongelmia aiheuttavat myös integraatiot ja näiden yhteydessä tai jälkeen tehtävät tiedon puhdistus, uudelleen duplikointi ja johdonmukaisuuden varmistustoimet, jotka hoidetaan joko räätälöidyllä logiikalla tai käsin tehtävillä korjausoperaatioilta.

Tietojen luotettavuuden luo tietojen jäljitettävyys lähdejärjestelmään, tiedolle tehtyjen käsittelyvaiheiden läpinäkyvyys mm. ETL-käsittelyt ja raporteille tehdyt rajaukset sekä kattavat kuvaustiedot ml. tietokuvaukset, koodistot, raporttikuvaukset, ja tietomallit.

Tietolähteen laadun arviointi- mitä huomioitava?

- Tietosuoja => tiukka suojaus voi heikentää aineiston analysointimahdollisuuksia
- Relevanssi => kuinka hyvin vastaavat tarpeeseen (esim. yrityksen liikevaihto vs. yrityksen myyty tuotanto)
- Tarkkuus ja luotettavuus => kuvaa täsmällisesti ao. ilmiötä
- Ajantasaisuus ja oikea-aikaisuus => kuinka uusia/vanhoja tiedot ovat
- Yhtenäisyys ja vertailukelpoisuus => suhteessa muihin tilastoihin, tilaston aikasarjoihin ja alueisiin
- Saatavuus ja selkeys => miten tiedot saatavissa (säännöllinen/epäsäännöllinen, kertaluonteinen, menetelmäkuvaus jne?)
- Tehokkuus => kuinka tehokkaasti tiedot tuotetaan prosessissa

(Storgårds, L. 11.6.2013, s. 11.)

3.7 Tietovaraston tietoturva

Tietovarastosta kertyy ajan myötä paljon tietoa, joka voi olla hyvinkin sensitiivistä ja siksi järjestelmä pitää suojata vahvasti tunkeutumisyrityksiltä. Samoin järjestelmän käyttäjien tunnistamisen ja käyttöoikeuksien hallinnan on oltava kunnossa. (Fujitsu Finland Oy 2015.) Tietovaraston tietojen suojaaminen on erityisen tärkeätä.

Tietovarasto on tietojärjestelmä muiden järjestelmien joukossa, eikä sen toteutukseen tai operointiin liity tietoriskejä, joita ei lähtökohtaisesti pystytä hallitsemaan ja hyväksymään. Kysymys on pikemminkin siitä mihin kohtaan valvonnan painopiste kannattaa siirtää tai on mahdollista siirtää ETL-tietovarasto-raportointiväline välillä. Painopisteen asettamisessa on löydettävä tasapaino riskien ja mahdollisuuksien välillä. Liian tiukat tietoturvan valvonta on omiaan vesittämään tietovarastolla tavoiteltavan hyödyn ja toisaalta liian lievä valvonta vaarantaa ennen kaikkea asiakkaiden tietosuojan.

Tietoturvatoiminnossa oleellisinta on se, mitä tietoja tietovarastosta näytetään käyttäjälle, eikä se mitä tietoja tietovarastoon viedään. Tämä tarkoittaa käytännössä sitä, että tietovarastoon voidaan viedä sama tieto kuin operatiivisessa lähdejärjestelmässä edellyttäen, että muutkin kuin tietoturvanäkökulmat on huomioitu.

Käyttöoikeudet tulee aina toteuttaa roolipohjaisesti ja niiden hallinta tulee integroida osaksi yrityksen käyttövaltuushallinta-prosessia.

Tietovarastosta on paljon erilaista tietoa saatavilla. Kaikki potilaiden kriittiset tiedot ovat helposti saatavilla sellaisessa muodossa, että niitä on helppo hakea ja käyttää. Tietoturvamääräyksien on katettava kaikki tiedot, jotka on poimittu tietovarastoon ja ajettu paikallisvarastoihin.

Käyttäjille myönnetään käyttöoikeuksia tietovarastokannan yksittäisten taulujen tai tietokannan näkymiin.

Yrityksen pitää määritellä tietovaraston turvallisuusvaatimukset tiedon hallinnan vaatimusmäärittelylle.

Tietoturvavaatimukset on määritelty vaatimusmäärittelyyn. Niiden pitää määritellä keillä on pääsy tietoihin ja keiden on voitava käyttää niitä.

Tyypillinen tietovaraston projekti käsittelee tietoja seuraavasti:

- Tiedot oleskelevat lähdejärjestelmissä ja valtuutetut käyttäjät pääsevät niihin käsiksi ;
- Tiedot on poimittu työtila-alueelle
- Tiedot ovat muuttuneet ja mahdollisesti ne on siirretty työtila-alueelle;
- Tiedot ladataan tietovarastoon ja
- Tiedot poimitaan tietovarastosta ja ladataan paikallisvarastoon, jossa loppukäyttäjät voivat käyttää sitä analysointia tai raportointia varten.

Jokaisen tietojen siirron prosessin vaiheeseen kuuluu tietty turvallisuusriski.

Vaatimusmäärittelyn tulee aina viitata lähdejärjestelmien tietojen käsittelyn turvallisuuspolitiikkaan.

Työtila-alue on ensimmäinen vaihei lähdetietojen poiminnalle ja tämä on paikka, jossa kaikki tiedot ovat käytettävissä lyhyen aikaa. Vaatimuksiin on määriteltävä käyttöoikeudet. Vaatimuksiin pitää myös sisältää tietojen hävittäminen siten, että arkaluontoiset tiedot suojataan sen jälkeen kun ne jättävät työtila-alueen.

Työtila-alueella voidaan myös tarvittaessa tallentaa tietoja, jotka hylätään laatuvaatimusten takia. Joskus nämä tiedot pysyy työtila-alueella, kunnes ne on korjattu uudella syötteellä ja joskus ne on saattanut korjata tiedon käsittelijä. Kaikissa tapauksissa turvallisuuden vaatimukset täytyy olla määritelty.

Käyttäjät tekevät Ad hoc tapauskohtaisia tietojen poimintoja suoraan tietovarastosta ja

toiset antaa vain hyväksytyn prosessin poimia tietoja. Käyttöoikeudet on määriteltävä.

Vaatimuksena voi olla esimerkiksi arkaluonteisten tietojen summaukset. Meillä on tietovarasto, joka sisältää lainsäädännön perusteella henkilötietoja. Päätämme yhdistää tiedot ulkoiseen tietolähteeseen, joka tunnistaa kaikki annetut lääkemääräykset tietyllä alueella. Vaikka lääkemääräysten tietoja ei yksilöidä lääkärillä tai potilaalla, saattaa olla mahdollista tunnistaa nämä henkilöt, jos alue on hyvin pieni esim. yksi lääkkeenmääräjä ja kaksi potilasta. Tällöin voimme määritellä vaatimuksen aggregaattitiedoista jos alue alittaa tietyn määrärajan.

Vaatimuksissa tulee määritellä, keillä käyttäjillä on pääsy tietoihin sekä aggregaattitietojen turvallisuussyistä johtuvat vaatimukset.

Huomioon otettavat muut turvallisuusnäkökohdat:

- Suunnittele vaatimuksia muille arkaluonteisille yksikön tiedoille esimerkiksi laskutuksen tiedot
- Suunnittele testaustietojen vaatimuksia - Onko tarvetta arkaluonteisille tiedoille?
- Dokumentoi turvallisuusnäkökohdat projektissa, esim. mitkä ovat vaatimukset luvattoman käytön estämiseksi
- Suunnittele määrittelydokumentaatiot niin, että niissä huomioidaan viittaukset tietohallinnon tietoturvapolitiikan yleisiin vaatimuksiin ja niihin on selkeästi määritelty projektiin liittyvät tietovaraston turvallisuus- ja testausvaatimukset.

Kaikkia tietovaraston tietoihin kohdistuvia riskejä ei koskaan voida täydellisesti poistaa. Yrityksen on kuitenkin tärkeää varautua tietovaraston erilaisiin käytön häiriöihin, mikä merkitsee tietoon liittyvien riskien kartoittamista ja niiden hallintaa.

Tietoturva tietovarastossa sisältää seuraavat tehtävät:

- Tiedon eheyden hallinta, taulujen constraints määritykset
- Tietomallit
- Taulujen lukitukset, päivitysten hallinta, kuten rivi / sivu lukitukset, duplikaatit
- Varmistukset / palautukset
- Lokitukset ja statistiikka tarkalla tasolla
- Kantaoikeudet, tekniset tunnukset, käyttäjä- ja ryhmätasoilla

3.8 Tiedon hallintamalli ja laadunvarmistusmenetelmä

Tiedon hallinnalla (data governance) tarkoitetaan hallintamallin prosesseja ja vastuita, joilla varmistetaan, että yrityksessä tuotettava tieto ja informaatio ovat luotettavaa ja hyödynnettävää, ja että tietoturva- ja tietosuojavaatimukset täyttyvät.

Tärkeintä tiedon hallinnan kehittämisessä on oman yrityksen liiketoiminnan tavoitteiden huomioiminen, joka luo tuottavuutta liiketoiminnalle.

Kuva 5. Data discovery ja tiedon visualisointi – enemmän irti tiedoista. (Helkiö, T. 2013).

Suurimmat näkemyserot liittyivät tietojen omistajuuteen (InfoBuild Oy 2010b). Monissa yrityksissä on vaikeata löytää tiedon omistajaa tai siitä on erimielisyyksiä.

Huhtala nosti esille yhtenä ratkaisuna näkemyseroihin vakuutusalalla yleistyvän tiedonhallintapolitiikan. Politiikka määrittää omistajan myös sellaisiin tietoihin, joiden omistajuus on yhteistä. Lisäksi siinä määritellään omistajan oikeudet ja velvollisuudet sekä tiedon laatutaso ja päivityssykli. (InfoBuild Oy 2010b.) Tiedon omistajuus on tietovaraston kannalta välttämätöntä selvittää.

Väärinkäsityksiä tiedonhallinnan kannalta saattaa aiheuttaa myös terminologian epäselvyys (InfoBuild Oy 2010b). Tiedon hallinnassa on syytä määrittää tieto ja varmistaa, että kaikki ymmärtävät sen samalla tavalla.

Päätöksenteko pohjautuu oikeaan tietoon. Tiedon laatua tarkkaillaan ja ylläpidetään jatkuvana prosessina.

Ulospäin tarjottavan paikallisvaraston (data mart) rakentamisessa on tärkeää huomioida tietoturva ja tietosuoja. Tästä syystä suositellaan fyysisten paikallisvarastojen rakentamista tähän tarkoitukseen tarkoitetulle tietokantapalvelimelle.

Tietovaraston laadunvarmistuksella on kolme tehtävää:

- Varmistaa tiedon laatu
- Ehkäistä virheiden synty
- Löytää syntyneet virheet mahdollisimman aikaisessa vaiheessa

Auditointi-taulut seuraa miten tietoja muutetaan ja se on hyvä tapa seurata tärkeiden tietojen muutoksia tai kun tietoja on näytetty. Auditointi-taulut seuraa, kun tietueita on katsottu tai päivitetty. Se näyttää kuka on katsonut tiedot ja sovelluksen, josta tieto on näytetty. (The Higher Ed CIO 2013.) Auditointi-tauluja pitää yrityksessä myös seurata.

Tietovarasto tarjoaa asiakkailleen korkeatasoista tietoa, ja tämän vuoksi laadunvarmistus on tietovarastoinnissa erittäin tärkeää. ISO-standardien mukaisesti laadukas tuote täyttää sille asetetut toiminnalliset ja ei-toiminnalliset vaatimukset.

Tietovaraston laadunvarmistuksen tavoitteena on varmistaa, että tietovarastossa oleva tietosisältö on riittävän laadukasta eli oikeaa ja ehjää palvellakseen tarkoitustaan.

Tietovaraston tiedon laatu muodostuu tietovaraston lähdejärjestelmien tiedon laadusta, tietovaraston sisällä tehtävästä tiedon yhtenäistämisestä ja yhteismitallistamisesta sekä laadukkaista kuvaustiedoista.

Tiedon laatuun kuuluvat tiedon eheys, kattavuus, yhtenäisyys ja aukottomuus.

- Tiedon johdonmukaisuus – tarkoittaa, että tiedot ovat keskinäisesti yhteensopivia ja tiedot ovat oikeita ennalta annettuihin ehtoihin nähden
- Tiedon kattavuus – kaikki tieto löytyy tietovarastosta
- Tiedon ajantasaisuus – tiedot toimitetaan ja ladataan tietovarastoon oikea-aikaisesti ja tieto on ajantasaista
- Tiedon saatavuus – joustavuus, tiedoilla saadaan täytettyä erilaiset tietotarpeet
- Master datan laatu

4 Johtopäätökset ja yhteenveto

Lähdejärjestelmän väärä tieto ei muutu oikeaksi, kun se ladataan tietovarastoon. Tiedon laatu koostuu monesta eri tekijästä. Tiedon laatua voidaan hoitaa automaattisesti ja osa on hoidettava manuaalisesti. Tiedon sisältö on sellainen esimerkki mitä ei voida aina hoitaa automaattisesti vaan jonkun ihmisen on se hoidettava. Joku voi kirjoittaa nimeksi Aku Ankka ja tämän tiedon voi korjata vain ihminen. Toisaalta väärän henkilötunnuksen saa automaattisesti.

Tiedon laadun arvo realisoituu usein yrityksessä vasta negatiivisessa mielessä, kun virheellinen tieto aiheuttaa suuria kustannuksia ko. yritykselle. Tietoa on hankala käyttää, jos siihen ei voi luottaa. Huonoa tietoa voi periytyä monesta eri lähdejärjestelmästä. Yrityksen johdolle täytyy kyetä osoittamaan laadukkaan tiedon tuottama arvo, joka voi olla lisääntynyt myynti, toimintojen tehostaminen tai riskien hallinta.

Opinnäytetyön kirjoittajalle tiedon laatu ja sen käsittely toi uusia näkökulmia itselleni. Näitä asioita voin sitten hyödyntää omassa työssäni tietovaraston asiantuntijana. Suurin yllätys oli se, kuinka monesta tekijästä tiedon laatu oikein koostuu.

Tiedon laatu vaatii jatkuvaa kehittämistä. Laadukas toiminta on systemaattista ja tähtää jatkuvaan parantamiseen. Tiedon hallintaa varten niitä tarvitsee mitata. Tietovarasto on tiedon arkistoinnin lisäksi aktiivinen päätöksentekoa ja päätösten toimeenpanoa tukeva ratkaisu. Tietovaraston tietojen laatuun ja sen koettuun arvoon vaikuttavat eri laatuominaisuudet.

Tietovarastoratkaisut tuovat operatiivisissa lähdejärjestelmissä sijaitsevaa tietoa. Mikäli tiedon syöttänyt ihminen tai automatisoitu prosessi mahdollistaa virheellisten tai tyhjien tietojen syötön ollaan asian ytimessä. Tietovarasto-hankkeissa pyritään varmistamaan tiedon laatu ja eheys hyödyntämällä siinä ETL-prosessia. Tavoitteena on saada tietovarastoon laadukasta tietoa, jota raportointi voi sitten hyödyntää. ETL-prosessia käytetään integroimaan eri lähdejärjestelmien tietoja yhteen.

Yllätyksiä tulee yleensä operatiivisten tietojen laadun tarkastamisen työläydestä. Väärä tai puutteellinen tieto tietovarastossa on aina katastrofi, jonka jälkeen käyttäjien luottamuksen takaisin saaminen on vaikeaa.

Tietovaraston rakentaminen parantaa yrityksen tiedonvaihtoa. Tietovarasto on se paikka yrityksessä, missä on paras tieto jokaisesta lähdejärjestelmästä.

Jatkotutkimuksena voisi selvittää, miten yrityksen tiedonhallintapolitiikka voitaisiin luoda ja kuinka ne pystyttäisiin jalkauttamaan.

Sanasto

Ad hoc –raportti	Ennalta määrittämätön raportti
Data Mart (DM)	Paikallisvarasto
Data Warehouse (DW)	Tietovarasto
Dynaaminen raportointi	Käyttäjän muokattavissa oleva raportti
ETL (Extract, transform and Load)	Tiedon poiminta, muuntaminen ja lataaminen
Master Data	Ydintieto on ydinkäsitteitä ja hitaasti muuttuvaa tietoa
Master Data Management (MDM)	Ydintiedonhallinta
Metadata	Metatieto on tietoa tiedosta, eli kuvailevaa ja määrittävää tietoa
Staattinen raportointi	Kiinteä raportti, joka ajetaan eräajona

Lähteet

Ari Hovi Oy 2012. Luettavissa:
http://www.arihovi.com/dataa-kaikille-kertaratkaisulla/
Luettu 4.10.2015

Ari Hovi Oy 2008. Luettavissa:
http://www.sytyke.org/lehtiarkisto/kirj/st19954/hovi954.htm
Luettu 7.10.2015

Eximia Business Intelligence Oy 2014. Luettavissa:
http://www.eximiabi.com/uusi-bi-ratkaisu-mista-aloittaa/?lang=fi
Luettu 6.10.2015

Fujitsu Finland Oy 2015. Luettavissa:
http://www.fujitsu.com/fi/solutions/business-technology/tietoturva/iot/
Luettu 7.10.2015

Helkiö, T. 2013. Affecto Oyj. Data discovery ja tiedon visualisointi – enemmän irti tiedoista. Luettavissa:
http://docplayer.fi/949174-Data-discovery-ja-tiedon-visualisointi.html
Luettu 11.10.2015

InfoBuild Oy 2010a. Luettavissa:
http://www.infobuild.fi/blogi/2010/08/tiedon-laadun-parantamisen-menetelm%C3%A4t
Luettu 7.10.2015

InfoBuild Oy 2010b. Luettavissa:
http://www.infobuild.fi/newsletter/2010/08/konsultti.html
Luettu 7.10.2015

Innofactor Oyj 2013. Luettavissa:
http://www.innofactor.fi/blogi/0/0/tiedon_laadun_jatkuva_parantaminen
Luettu 7.10.2015

Louhia Consulting Oy 2015. Luettavissa:
http://www.louhia.fi/artikkelit/tietovarastot-ja-raportointi/
Luettu 7.10.2015

Miracle Finland Oy 2015. Luettavissa:
http://www.miracleoy.fi/palvelut/konsultointipalvelut/tietovarastointi/tietovarastolatausten-dw-suunnittelu-ja-toteutus/
Luettu 7.10.2015

Midagon Oy 2011. Luettavissa:
http://www.midagon.fi/blog/mika-niissa-luotettavissa-raporteissa-oikein-kestaa/
Luettu 7.10.2015

Niemi, K. 18.4.2013. Salcom Solutions. Salcom Group Oy. Data-suomi-sanakirja. Termi-viidakon selviytymisopas. Luettavissa:
http://www.slideshare.net/kalleniemi1/data-suomi
Luettu 12.10.2015

Octel Oy 2008. Luettavissa:

http://www.pcuf.fi/sytyke/lehti/kirj/st19954/stahl954.htm
Luettu 7.10.2015

Oksanen, M. 12.04.2011. Avarea Oy. Liiketoimintatiedon hallinta ja hyödyntäminen (BI). Luettavissa:
http://www.nicetuesday.fi/Liiketoimintatiedon%20hallinta%20(BI).pdf
Luettu 7.10.2015

Rongo Oy 2014. Luettavissa:
http://www.rongo.fi/2013/10/perustietojen-hallinta-yksi-totuus-tiedolle/
Luettu 4.10.2015

Storberg, P. 2015. Business Intelligence, Data Warehouse ja ETL. Luettavissa:
http://pate.kapsi.fi/port/13.php
Luettu 11.10.2015

Storgårds, L. 11.6.2013. Tilastokeskus. Tilastotiedot yhteiskunnan muutosten ja kriisien kuvaajana. Luettavissa
http://www.stat.fi/ajk/tapahtumia/2013-06-11_tampere_storgards.pdf.
Luettu: 12.10.2105

Talmax Oy 2013. Luettavissa:
http://www.talmax.fi/tietovarastot-ja-integrointi
Luettu 5.10.2015

The Higher Ed CIO 2013. Luettavissa:
http://blog.thehigheredcio.com/2013/04/10/data-warehouse/
Luettu 7.10.2015

Theta 2015. Luettavissa:
https://www.theta.co.nz/solutions/business-intelligence/business-intelligence-and-data-warehouse-methodologies
Luettu 5.10.2015

Teknologian tutkimuskeskus VTT Oy 2015. Luettavissa:
http://www.vtt.fi/sites/hti/mit%C3%A4-k%C3%A4ytett%C3%A4vyys-tarkoittaa
Luettu 7.10.2015